



City Research Online

City, University of London Institutional Repository

Citation: Bastos, M. T. (2015). Shares, Pins, and Tweets: News readership from daily papers to social media. *Journalism Studies*, 16(3), pp. 305-325. doi: 10.1080/1461670X.2014.891857

This is the accepted version of the paper.

This version of the publication may differ from the final published version.

Permanent repository link: <https://openaccess.city.ac.uk/id/eprint/17081/>

Link to published version: <https://doi.org/10.1080/1461670X.2014.891857>

Copyright: City Research Online aims to make research outputs of City, University of London available to a wider audience. Copyright and Moral Rights remain with the author(s) and/or copyright holders. URLs from City Research Online may be freely distributed and linked to.

Reuse: Copies of full items can be used for personal research or study, educational, or not-for-profit purposes without prior permission or charge. Provided that the authors, title and full bibliographic details are credited, a hyperlink and/or URL is given for the original metadata page and the content is not changed in any way.

Shares, Pins, And Tweets: News Readership from Daily Papers to Social Media

Marco Toledo Bastos

Accepted for publication in Journalism Studies
(pre-publication version: some changes still possible)

This paper compares the volume of news articles per section in newspapers and social media platforms. To this end, two weeks of news articles were retrieved by querying the public Application Programming Interfaces (APIs) of The New York Times and The Guardian and the diffusion of each article on social media platforms Twitter, Facebook, Google+, Delicious, Pinterest, and StumbleUpon, was tracked. The results show significant differences in the topics emphasized by newspaper editors and social media users. While users of social media platforms favor opinion pieces, along with national, local, and world news, in sharp contrast the decision of news editors emphasized sports and the economy, but also entertainment and celebrity news. Common to social networking sites is the prevalence of items about arts, technology, and opinion pieces. Niche social networks like StumbleUpon and Delicious presented a greater volume of articles about science and technology, while Pinterest is mostly dedicated to fashion, arts, lifestyle, and entertainment. Twitter is the only social network to have presented a statistically significant correlation with the distribution of news items per section by The Guardian and The New York Times. The results of this study provide a bridge between journalism and audience research and present evidence of the differences between readership in social and legacy media.

1. Introduction

The early literature on online news established that online journalism often failed to explore interactive features and mostly relied on content from the parent publication. Newsmaking decisions were driven by print newsrooms and articles for the online version rarely included external links or multimedia content. Quandt (2008) analyzed the impact of the internet on newspapers by a comparative content analysis of 10 online news media in four European countries and the United States. Despite the avenues of opportunities provided by the internet in page design and layout, and also in the potential for exploring new formats for writing and telling stories, Quandt findings were restricted to national differences reflecting particular journalistic cultures and audience interests. The material from online portals matched the content exhibited in the print version, with a lack of multimedia features, absence of options for contacting journalists, and missing attributions to news sources.

These results confirmed previous investigation on the influence of print to online editions, which mostly consisted of copying the printed newspaper (Neuberger et al. 1998). But despite the slow pace of editorial changes in online journalism from the late nineties to the late noughties (Greer and Mensing 2006), newsrooms at large became gradually more dependent on networking technology, with a number of news outlets appointing social media editors and integrating print and digital operations. These structural changes in the newsroom reflected a broader trend in the early 21st century that shifted newspapers towards compact, visually-appealing, and commuter-friendly editions featuring fewer stories with fewer words

on each page and making space for the use of photography in the storytelling, especially on the front page (Franklin 2008).

During this critical decade, news editors tried to meet the increasing demand for articles dedicated to sports and entertainment and allowed for the growing number of gossip news about Hollywood celebrities and advice columns. Franklin (2008) commented that hard news stories disappeared from the front page, which simply featured large photographs and headlines, and that a number of editorial changes were introduced to emphasize more personalized news, with a focus on themes such as education, the environment, health and housing, along with an emphasis on readers' letters and readers' forum. The dominance of celebrity and social news, and the increasing growth of reality shows and other forms of popular-culture oriented news, contributed to the blurring of credibility boundaries that once set traditional outlets apart from digital media (Johnson and Kayer 2004).

The integration of print and online newsrooms also reflected the reality of newsmaking in face of a multitude of media channels. Schudson (2011) enumerated the decisive factors in the overall change in the news ecosystem, which stems from the collaboration between reader and writer: the lack of ultimate distinctions among tweets, blog posts, newspaper stories, magazine articles or books, and the diminishing gap between professionals and amateurs. Schudson's conclusion was that the line between old media and new media had been blurred beyond recognition and that the very nature of news values was evolving. By incorporating interactive components of digital platforms, newspapers initiated a process of changing the editorial decisions that define the newsmaking process.

The year of 2012 marked a decisive departure from the previous decade with American news outlets taking a decisive turn towards digital news. When Neuberger et al. (1998) interviewed readers from 81 daily newspapers in Germany that housed a digital version of the paper, they found that two-thirds of users would have chosen the print newspaper against the online version if only one version was available, and that in this scenario less than a third of the audience would have preferred the online version of the newspapers. Fourteen years later the percentage of American adults younger than 30 who read news on social networking sites was equal to television (34%), with just 13% reading news on newspapers (Pew Research Center 2012a).

Pew's report indicated that the proportion of Americans reading news on a printed page declined to 23% down from 26% in 2010. Most regular readers of The New York Times declared they read the paper mostly on a computer or mobile device (55%), as did 48% of regular USA Today and 44% of Wall Street Journal readers. The figures in Britain showed a similar trend towards online readership, with just over half (55%) of internet users reading newspapers or news service online in 2011. In contrast to the US, where the complementarity between offline and online readership was declining, offline reading was still highly important in Britain. The percentage of Britons reading newspapers only offline in 2001 was around 30%, while the percentage reading both online and offline was about 60% (Dutton and Blank 2011).

Despite the sizeable literature on the declining readership of newspapers (Atanassoff and Oehlmann 1997; George 2008), research that evaluates the impact of social networks to newsmaking is still forthcoming (Newman 2009; Lasorsa et al. 2011; Bastos et al. 2012; Aitamurto and Lewis 2013). To fill this gap, this paper compares the editorial decisions of news editors regarding the allocation of news items per section with the feedback from social networking sites to the news articles. For the purposes of this investigation, we focused on American daily The New York Times and British newspaper The Guardian — national newspapers with international editions, global internet presence, and in-depth coverage of international affairs. The Times and the Guardian are also the object of a sizable literature comparing the coverage of both papers (Hopple 1982; Bantimaroudis and Ban 2001).

In order to compare the volume of articles per news section in newspapers and social media, we queried the public APIs of The New York Times and The Guardian and extracted all news articles published during the first two weeks of October 2012. We subsequently mined six of the largest social bookmarking and social networks in the United Kingdom (comScore 2013a) and the United States (comScore 2013b) to measure the impact of each news article in each social media platform. The social networking sites investigated are: Twitter, Facebook, Google+, Delicious, Pinterest, and StumbleUpon. These services serve different goals and rely on distinct user bases, but they have in common the potential for sharing news items.

While Twitter aims to offer a wire-like service for news organizations (Cowan 2010) and Facebook presents very large cascades (Dow et al. 2013), Delicious, StumbleUpon, and to some extent Pinterest are dedicated to social bookmarking and archiving internet content beyond the scope of news. While Facebook and Twitter are primary representatives of social networking sites, StumbleUpon and Delicious are social bookmarking sites with recommendation systems that incorporated social features and photo-sharing capabilities similar to Pinterest, a pinboard-style website that allows users to manage theme-based image collections. Google+ relies on Google search referrals and offers the potential to bridge social bookmarking and social networking sites. We expect this comprehensive set of social networks to provide a reliable representation of the social media ecosystem.

The remainder of the article is organized as follows. In the next section of the paper we review the literature and in section 3 we state the objectives of this study. In sections 4 and 5 we describe the methods and the data used in this investigation. The last two sections of this paper provide an overview of the results and a discussion about the findings.

2. Previous Work

Hart (1966) presented the first study to compare the percentages of allotted space to world news in four English and American newspapers by reading and measuring the space dedicated to world news in editorials, signed columns of news comment, analysis and letters to the editors, editorial cartoons, and pictures with cutlines. Hart found that English newspapers judge foreign news as more newsworthy than their American counterparts, although all newspapers investigated agreed as to what type of news was important to print. Hart also found that of the eight newspapers in the study, only one English newspaper, The Guardian, and one American newspaper, The New York Times, printed as much as 80% of its foreign coverage as news stories, with most newspapers in the study covering world news with pictures, editorials, and cartoons.

The first study to explore the relationship between newspaper's editorial profile and audience feedback was carried out by Neuberger et al. (1998). The authors investigated 81 daily newspapers in Germany that had a web presence in 1997 and described variations between print and online editions. By comparing the popularity of each news section of the newspaper with the editorial profile of the outlet, the authors found that traditional news sections are the most frequently read, with niche sections like science and technology also being very popular. The authors found clear differences between nationwide and local newspapers, with national newspapers reporting more politics, economy, arts, and science, and less local news, ads, entertainment, and quizzes. The research also found that the preferences of print media are transferred to the online version of the newspaper, with readers of regional newspapers reading local news and readers of national dailies engaging to up-to-date information on politics and business matters.

Kamerer and Bressers (1998) analyzed the content of 105 online U.S. dailies for the presence of national and local news and reported an increase in the percentage of sites featuring national news and a prevalence of local news across websites. Dans (2000)

compared the patterns of news consumption and the levels of readership online and offline in 15 Spanish newspapers. The researcher found that the reading patterns of online news strongly differed from those of print newspapers, with online readership being more goal oriented — particularly in the comparison between weekday and weekend circulation. Althaus and Tewksbury (2002) investigated whether readers of print and online versions of The New York Times reported different perceptions of political issues. The researchers found that readers of the print edition selected news based on how the sections were presented and picked topics that appeared on the front page and in the first section, while readers of the online edition selected topics in a less determined fashion and gave preference to a wider range of topics in comparison to readers of the print edition.

D’Haenens et al. (2004) investigated variations in readership between digital and print versions of two Dutch newspapers and reported no significant differences in the consumption of news in print and online platforms. However, the researchers noted a significant difference in the amount of time spent reading each section of the print and online newspaper and found that more time was spent reading the online versions of the newspapers across all news sections, except for sports and local news. D’Haenens et al. (2004) also found that online readers recalled more international news and less sports than readers of the print version, which is consistent with the much higher volume of local news and sports in the print edition of the newspapers. The research showed substantial asymmetries between online and offline readership, particularly in regard to sports and local news items.

Gasher and Gabriele (2004) investigated whether online publishing has prompted the Canadian daily newspaper Gazette to change its news coverage. The study reported that the online version of the paper consistently carried more international news items than its hard-copy edition, a distinction likely due to the website’s reliance on wire-service copy. Quandt (2008) compared the distribution of articles per news sections in online news websites with print press and TV coverage. The researcher found a similar pattern across the news ecosystem, with an emphasis on national politics and economy, followed by human-interest stories, international politics, crimes, sports, and culture. The investigation reported that online journalism was fairly conventional in regard to the distribution of news items per news section (or topic categories). The research reported striking similarities between websites inside a country, but also remarkable differences from country to country.

3. Objectives

The organization of news articles according to news sections is a legacy from print media that both classified information according to topics and ensured that a diverse selection of content was covered by the press. The dominant topics of news coverage depend on editorial decisions made by newspapers and newsmen that are bound to be affected by audience agency in social media. Therefore, the primary objective of this investigation is to answer the following questions: How do newspapers compare to social media in the ratio of news items per news section published and shared? Which news sections receive more coverage and which are the mostly shared topics by social media users? How much overlap there is between the editorial choices of newspapers that decide which topics receive more attention and social media feedback in terms of shared, tweeted, pinned and clicked news articles? Finally, we evaluate the claims that Twitter is a prime venue for news in the social media ecosystem (Cowan 2010; Wolff 2013).

Therefore, this study contrasts what newspaper editors had deemed to be of public interest and/or the interest of the public with audience feedback to the same news items. To this end, we compare the distribution of news articles on news websites with the volume of articles shared on Facebook, retweeted on Twitter, pinned on Pinterest, or recommended on Google+, Delicious, and StumbleUpon. The present study has thus two secondary purposes.

The first is to establish the relative distribution of hard, soft, and general news (Lehman-Wilzig and Seletzky 2010) crafted by the editors of The New York Times and The Guardian. The second is to evaluate how the distribution of hard, soft, and general stories in newspapers compares to the largest social media platforms (Facebook and Twitter).

The factors governing the newsworthiness of information were originally defined by Otto Groth (1928) and included seven newspaper qualifications and a number of article attributes. Galtung and Ruge (1965) further explored these categories and identified thirteen factors tested against the hypotheses of additivity, complementarity, and exclusion. These principles were again revisited by Tunstall (1971) who showed that because Galtung and Ruge's data was restricted to the coverage of international crises, they ignored day-to-day coverage of lesser, domestic, and mundane news. The reevaluation of Galtung and Ruge's seminal investigation led to further research on the factors driving newsworthiness and to a general consensus that the context of print media was of increasing editorial emphasis on entertainment (Franklin, 1997).

The emphasis on soft news items allowed for coverage of events directed by marketing departments and heavily influenced by audience share. Harcup and O'Neill (2001) commented on Franklin's work and pointed out that no contemporary set of news values could be complete without the entertainment factor. The authors revised Galtung and Ruge's original set of factors to include Entertainment and Good News (in opposition to Bad News) and to highlight the shifting paradigm of newsworthiness from mid-twentieth-century news reporting, focused on political and socio-economic issues, to infotainment news covering celebrities' personal lives and showbiz events. Other studies have investigated the proportion of soft and hard news in television (Scott and Gobetz 1992), but the concept of newsworthiness remained untested since the onset of social networks. We expect the comparison between the distribution of news items on legacy and social media to contribute to the understanding of newsworthiness according to news editors and social media users.

4. Methods

Since the 1950's a number of studies have investigated journalism balance and news coverage on newspapers and television using content analysis to show how newspapers provided more coverage to specific events (Klein and Maccoby 1954). This method was consistently deployed in the following decades by Hart (1966), McCombs and Shaw (1972), Koszyk and Prause (1987), and Kamerer and Bressers (1998) — likely due to the technical challenges involved in retrieving and analyzing large-scale, quantitative data from different news outlets. However, the recent availability of online news articles and social network data, which can be retrieved from standardized, public APIs, provided avenues for quantitative research at unprecedented scales.

We used tools designed to query newspapers APIs (Bastos and Puschmann 2013) and extracted all news articles published between Monday, 1st of October 2012 (00:00:00 GMT) and Sunday, 14th of October 2012 (23:59:59 GMT), thus including two weeks of data collection. We used a list of 200 frequent keywords from each newspaper's front page from the 1st to the 14th of October and a list of 22 newspaper sections to continue querying the APIs. We also searched the APIs for the full content via global search. After collating the data, The New York Times API returned 11,607 and The Guardian 5,222 unique news articles. These figures are consistent with the number of news articles published daily by each news outlet, as The New York Times publishes approximately 1,000 items per day (Caruso 2012) and The Guardian around 350 articles of content daily (Hume 2013).

We compiled a list with the URLs of each news article retrieved from the newspapers and subsequently queried the APIs of Twitter, Facebook, Google+, Delicious, Pinterest, and StumbleUpon for the number of hits of each news item published by The New York Times

and The Guardian in the period. We also queried the APIs of LinkedIn and Digg, but the platforms did not return substantial number of hits. Pinterest hits are highly concentrated in three news sections, but we decided to keep the data because it is statistically significant. To evaluate the variance between the distribution of items in legacy and social media, we compared the sheer volume of articles per section instead of the length of the articles.

4.1 Particulars of the Method

We analyze the difference between the volume of items and the length of articles per news section and found a much smaller variation than reported in previous investigations based on a sample of news items (Hart 1966), with major news sections (politics, economy, arts, opinion, national, and local news) presenting equivalent variance between number of items and number of words per item. News articles about fashion, entertainment, and lifestyle are on average shorter on both newspapers, while items on education, politics, and sports are slightly lengthier. For the purposes of this investigation, we considered the distribution of news items per section instead of word count.

The APIs of Times and Guardian also return metadata related to the news items, including the news sections the items were assigned to by the editors. We relied on this information to investigate the relative predominance of news sections across legacy and social media. To this end, we focused on 19 news sections that could be aggregated based on the data retrieved from the APIs and that are equally distributed across hard, soft, and general news. These sections are: arts, cars, economy, education, entertainment, environment, fashion, health, jobs, lifestyle, local news, national news, opinion, politics, science, sports, technology, tourism, and world news. We computed the number of items and the number of words per news section. Twitter streaming API allows for the continuous tracking of news items published in the platform, so we tracked the spread of news items on Twitter in order to compare the lag between the time news items go online and the moment they are tweeted. Lastly, we calculated the statistically significant correlations between the distribution of news items per section in newspapers and social media.

4.2 Limitations of the Method

Some potential limitations of this study are related to the availability of data, particularly regarding the ratio of news articles made available through public APIs. Although REST APIs are customizable and reliable sources to access news articles (particularly in comparison to webpage scrapping), we cannot determine whether media outlets have made available the totality of their content via API. To control for the data returned from each API, we queried 200 keywords and compared the aggregated results with queries without keywords. The results indicated that queries without keywords returned a higher volume of news articles and also included the articles in the control sample.

In order to control for issues related to Tweet Button (Twitter API Documentation 2012), we calculated retweet count by calling Twitter REST API at the end of the period and by keeping track of the counting via Twitter Streaming API, which imposes challenges due to the API returning URL links in a shortened format. In order to match the results from the media outlet APIs and the results from Twitter Streaming API, we used a three-pass routine to resolve nested shortened URLs (i.e. previously shortened bit.ly URLs that are again shortened by Twitter t.co URL wrapper) and to retrieve the target URLs. Facebook only provides access to the REST API, which returns the number of shares of each link by calling Facebook Graph API.

There is also a limitation related to the time when news articles are posted by news outlets and the period during which social media users read, tweet, share, pin, and post links

to the news items. We found that news articles continue to circulate on social networks on average two days after they are published by the news outlets. Because we defined the time frame of this study to include the first two weeks of October, we expect news articles published during the last days of the second weekend of October (October 13-14) to present a lower-than-average number of hits on social media in comparison to the previous days.

Moreover, social media feedback to articles may depend on the appeal of the article rather than the topic in which the content was classified. Lastly, the classification of news articles in news section is a decision made by newsroom editors that is not necessarily consistent across different newspapers. Tabloid newspapers tend to include a large number of soft news items in the economy section, particularly items on consumer products and tourism, whereas such items generally appear in the lifestyle section of quality papers. Political news can also be arbitrarily assigned to world, national, or local news depending on editorial decisions that are particular to each newspaper. Although we cannot determine the consistency of the criteria used to assign items to news sections in the decision-making process of each newspaper, this can be partially controlled for by analyzing only quality papers instead of tabloid and popular papers, thus relying on a more uniform baseline of news editors' decisions.

5. Data

The data was collected over a two-weeks period and amounts to approximately 5,000 news articles published by The Guardian and 11,000 by The New York Times. Consistent with the volume of news articles published by each outlet daily, Times' API returned 83% and Guardian's 106% of the expected number of news items. The first article published by The Guardian is dated the 1st of October at 00:09 and the last item the 14th of October at 23:47. The New York Times released the first article at 00:00 on the 1st of October and the last at 23:52 on the 14th of October. During the period of analysis we tracked the diffusion of each news article on the social networking sites Twitter, Facebook, Google+, Delicious, Pinterest, and StumbleUpon. This method allowed for identifying the number of articles published in each news section of the papers and their relative impact on social media platforms without directly resorting to content analysis.

In the remaining sections of this paper we focus mostly on Facebook and Twitter because these are currently the largest social networking sites worldwide with 1 billion and half a billion users respectively (Whittaker 2012; Shiels 2011), but we also analyze the results from four additional social networks (Pinterest, Google+, Delicious, and StumbleUpon) to evaluate the impact on social media of items published by the Times and the Guardian. The upper part of Table 1 shows that over one-third of the news items published in the period by The New York Times showed up on Twitter stream of public tweets through 167,189 retweets (39 retweets per news article on average). Even more links appeared on Facebook — 40% as opposed to 36% on Twitter — with over 2 million users sharing items published by the Times on Facebook (445 shares per article on average).

Guardian's presence on social media is also substantial. Most of the content published by The Guardian in the period (86%) appeared on Facebook stream via 856,149 shared links (190 shares per news article on average). Guardian's presence on Twitter is yet more significant — 96% as opposed to 86% on Facebook —, with over a quarter million retweets (50 retweets per news article on average) during the first two weeks of October. The lower part of Table 1 shows that The Guardian has considerably more presence on the remaining social networking sites. Even though the Guardian published less than half of the volume of articles published by The New York Times in the period, it harnessed twice as many hits on Google+, two-and-a-half as many hits on StumbleUpon, and considerably more hits on Pinterest and Delicious (23% and 43%, respectively).

[INSERT TABLE 1 HERE]

There is also substantial crossover in the location of social media users that tweeted articles to the two papers, as indicated by the sample of tweets with links to The New York Times and The Guardian that included geolocation (Figure 1). The APIs of the Times and the Guardian return rich information about each news article, including the news sections, word count, author, date of publishing, and URL. We examined the data to identify news articles supplied without information about the news section and/or word count, and for these articles we proceed as follows: word count was calculated based on the full text of news items that lacked word count; news sections were assigned by exploiting the format of links created by the news outlets and taking advantage that links consistently display the news section in which they appear (e.g. world news, entertainment, sports).

[INSERT FIGURE 1 HERE]

6. Results

The chronological distribution of articles by the Times and the Guardian is fairly symmetrical (Figure 2), with the former releasing more content during the middle of the week and the later pushing a proportionally higher volume of news articles on Fridays. Figure 2 shows that news output plummets on weekends in the two news outlets, even though print circulation of the Times is higher on weekends than weekdays. The Guardian does not circulate on Sundays, when the sister paper The Observer takes over the newsroom. However, Guardian's API returns news items from The Guardian and The Observer, so the results are not affected by Guardian having no issues on Sundays. Figure 2 shows a time series of articles published by both outlets, including items from Guardian's sibling paper The Observer.

[INSERT FIGURE 2 HERE]

We found that Twitter is the only social network to present a statistically significant correlation with the distribution of news items per news sections on the Guardian and the Times ($r=0.88$ and $r=0.68$, respectively, $p<0.001$). This is indicative of Twitter as a preferred channel to stream a wide variety of news content and a good indication that Twitter users feed the network with a set of items as varied as those in news outlets. Table 2 shows the statistical correlations between the distribution of news articles per section on papers and on the social networking sites investigated in this study (only significant correlations are shown, $p<0.001$). Pinterest, Google+, and Stumbleupon reported high correlations but non-significant P values for both newspapers, and therefore are not included in the table. Facebook and Delicious presented statistically significant correlations only with the distribution of news items published by The Guardian.

[INSERT TABLE 2 HERE]

We further explored how closely the distribution of news items on Twitter follows news outlets. News articles not only reproduced the diversity of topics covered by the papers (Figure 3b and Table 2); they are also tweeted shortly after being published by the outlets. Figure 3a shows a time-series of the interval between news outlets publishing the articles and users posting links to the news items on Twitter public stream. The plots on the right side show a fitted linear regression between the distribution of news items per section on the

papers and on Twitter. We understand these results to show a linear relationship between the editorial profile of online newspapers and Twitter, mostly due to the growing popularity of Twitter among journalists, with most journalists in the United Kingdom relying on Twitter to source or piece together information (Broadgate Mainland 2011).

[INSERT FIGURE 3 HERE]

The comparison between Twitter and Facebook as news sources highlights Facebook emphasis on items about arts, but mostly on opinion pieces, which account for nearly one-third of all The New York Times articles shared on Facebook. Twitter, on the other hand, presents a higher percentage of items about economy, technology, and national news. In fact, just under a quarter of all messages from Twitter public stream with links to Times articles are dedicated to national news (as opposed to 9% on Facebook). Figure 4 shows the breakdown of the data organized by social networking sites and presents the relative number of news items per section. The bar chart should be read horizontally — bars of the same color sum up to 100%.

[INSERT FIGURE 4 HERE]

These figures show the contrast between the two major social media platforms as news sources. The differences are further accentuated when we take into account the original volume of news articles published in each newspapers' section. Figure 5 presents a comparison between Twitter and Facebook and also shows the relative numbers of articles published in each news section. The sections emphasized by the newspapers are considerably similar, with most articles on the Times and the Guardian covering world news (20% and 12%, respectively) and sports (16% and 15%, respectively). The newspapers also presented a proportionate distribution of items across niche sections like arts (6% and 7%, respectively), health and cars (1% and 0%, respectively, for both sections), science (1% and 2%, respectively), tourism (0% and 1%, respectively), fashion (1% for both papers), technology (2% for both papers), and politics (5% for both papers).

But the differences in emphasis between the two outlets are also significant. While the Times hardly published items about entertainment (2%), the topic harnessed 17% of the total volume of articles published in the period by the Guardian. The New York Times emphasizes hard news topics, particularly in comparison to The Guardian, such as economy (20% and 8%, respectively), local news (3% and 1%, respectively), national news (16% and 7%, respectively), and world news (20% and 12%, respectively). The Guardian, on the other hand, run comparatively more opinion pieces (8% against 3% in the Times), education and jobs (2% against 0% in the Times, for both sections), entertainment (17% against 2% in the Times), lifestyle (4% against 2% in the Times), and environment (4% against 0% in the Times). Common to both newspapers is the large volume of articles about sports that fail to actively engage social media users.

[INSERT FIGURE 5 HERE]

The results show that social media users express a preference for a subset of content and information that is at odds with the decisions of newspapers editors regarding which topic to emphasize. Only a quarter of articles about sports published by the Times ever showed up on Twitter or Facebook; only about half of sport-related items published by The Guardian appeared on Twitter, and only one-third on Facebook. World news articles are published by the editors of the Times (20%) more often than by users of Facebook (10%) or Twitter (10%), while the opposite can be observed for The Guardian. Only 12% of the content published by

the British outlet is dedicated to world news, but these few links are responsible for 16% of Twitter stream and 19% of Facebook stream with links to The Guardian.

The results also show that social media users that read the publications tend to favor hard news over soft and general news in a proportionally higher ratio than do newspaper editors. Figure 6 presents a comparison between the distribution of news articles by soft, general, and hard news in the newspapers and on Twitter and Facebook. As commented in the previous example, the emphasis placed by newspapers on sport-related items is in sharp contrast with the feedback from social media users. Although most news sections are uniformly and symmetrically distributed across newspapers and social networking sites, we found remarkable differences on the number of news items about arts, science, technology, and opinion pieces, which are on average more frequent on social networking sites than on newspapers. The Times also placed greater emphasis on items about economy, particularly in comparison to social media, while the Guardian published significantly fewer articles in the economy section.

[INSERT FIGURE 6 HERE]

Lastly, we compared the distribution of news items per section in newspapers with the relative number of hits across the six social networking sites. We found that the distribution of articles per news sections on social media depends to a certain extent on content decisions of news editors — i.e. The New York Time's skimpy coverage on environment-related topics is followed by a flatline performance of these articles on social networking sites (less than 1% of total hits). Guardian's emphasis on environmental issues is also replicated on social media. Environment-related pieces reached comparable distribution shares across the six social networking sites, with a higher-than-average distribution on Twitter and Facebook and a lower-than-average distribution on Delicious, Pinterest, and StumbleUpon.

[INSERT FIGURE 7 HERE]

Figure 7 shows the percentage of items per news section on the Times and the Guardian compared to the percentage of hits per news section on social networking sites Twitter, Facebook, Google+, Delicious, Pinterest, and StumbleUpon. Pinterest is a networking site that works like a bulletin board for photos and is not usually associated with news. Pinterest, although primarily a visual medium, shows considerable traffic to news articles and generated on average half of Google+ hits (4,390 against 11,532 for the Guardian and 3,561 against 5,626 for the Times). Pinterest's 4,390 hits for Guardian articles resulted from a total of 5,222 articles, which averages nearly one hit per article published by the Guardian. Although these results are relatively expected, they also show the importance of each social network to niche sections. Pinterest is mostly dedicated to fashion, arts, entertainment, and lifestyle — sections that account for more than half of all articles that showed up on Pinterest stream (and in sharp contrast with the relatively low emphasis placed on these topics by news outlets and the remaining social networks).

StumbleUpon, a recommendation engine similar to Digg, shows a much higher-than-average volume of news items about tourism, science, entertainment, and arts. Delicious, a social bookmarking tool that allows users to access their favorite websites from any computer, presents a higher-than-average percentage of items about technology, tourism, and economy. But economy, by all means, remains a domain of news outlets, with nearly all social media platforms streaming fewer articles on economy in comparison to the press coverage. Google+ shows a prevalence of articles about local and word news, technology, science, and opinion pieces. Facebook presents a much higher-than-average volume of items dedicated to arts,

environment, opinion pieces, and perhaps surprisingly, world news. After Pinterest, Facebook is also a prime venue for content about fashion and entertainment.

Figure 7 also shows that Twitter is a platform more centered on hard news stories. The social networking service based on 140-characters messages streams on average more content about national and world news, politics, environment, education, and economy, often streaming a volume of hard news stories not only higher than the other social networking sites investigated in this study, but also higher than the average distribution found in the news outlet that provides the content. In comparison to the daily editions of The New York Times and The Guardian, Twitter stream emphasizes arts, education, environment, opinion pieces, politics, science, and technology.

7. Discussion

We have presented a number of comparisons between the topics emphasized by media outlets and by social media users, but we also analyzed the differences between the two newspapers. In comparison to the Guardian, the Times emphasizes economy, local, national, and world news, while the Guardian provides more opinion pieces and articles about education, entertainment, lifestyle, and environment. Common to both newspapers is the large volume of articles published about sports that fail to actively engage the social media audience. We found that economy is a domain of news outlets, with nearly all social media platforms streaming fewer articles on economy in comparison to the press coverage.

Although we confirmed that the distribution of news articles on social media platforms depends to a certain extent on the decisions of news managers (including news editors, assignment editors, and section editors), we also found the proportion of new articles per section posted on social networking sites to be at odds with the allocated space for news articles on the online version of the newspaper. This is in contrast to similar studies that have compared print and online newspapers and found that the preferences of print media are transferred to the online version of the newspaper (Neuberger et al. 1998). Therefore, we expect the results of this study to provide compelling evidence of the differences between readership in social and legacy media.

We found that The Guardian harnessed more hits on social networks than The New York Times proportional to the volume of articles published in the period. It is our understanding that these figures result from Guardian's strategy of providing online news free of charge, while the Times has introduced online paywall systems in the period covered in this study. We also found that Twitter is the only social network to present statistically significant correlations with the distribution of news items per section on The Guardian and The New York Times ($r=0.88$ and $r=0.68$, respectively, $p<0.001$). This is indicative of Twitter as a primary channel to stream a wide variety of news content, and a good indication that Twitter reproduces the diversity of topics covered by newspapers. The comparison between Twitter and Facebook as news sources shows that Facebook emphasizes arts and opinion pieces, while Twitter presents more content dedicated to economy, technology, and national news.

Common to the social networks investigated in this study is the prevalence of items about arts, technology, and opinion pieces. Niche social networks like StumbleUpon and Delicious presented a greater volume of articles about science and technology, while Pinterest is mostly dedicated to fashion, arts, entertainment, and lifestyle. After Pinterest, Facebook is also a prime venue for content about fashion and entertainment. Twitter is the social networking site with the higher ratio of information centered on hard news stories. In comparative proportions, the social networking service based on 140-characters messages streams more items about national and world news, politics, environment, education, and

economy than any of the other social networks studied in this paper, often streaming a higher volume of hard news stories than the average distribution found in the content provider.

The transition from the print press to online news websites represented a qualitative leap that changed readership due to decisions relating to the platform design, section assignment, and website layout. These technical modifications exert a major influence on reading habits and change the way news is consumed. The transition from online to social networking news alters the reading experience not only because of changes in visual design, illustrations, typography and positioning, but also due to the technical possibilities granted to audiences that can access, share, and recommend the content to large communities of users. In fact, the impact of social networking sites to the news industry is likely to be of greater importance than that measured in this study.

Moreover, the differences between what newspaper editors have deemed to be of public interest and what is of interest to users of social networking sites needs be considered within the wider debate on press and the public sphere and the validity of liberal journalism's claims to freedom and objectivity (Singer 2007; Christians et al. 2009). As readership agency begins to deliver critical feedback to news items and interfere in the agenda of legacy media, newsrooms will have to strike a balance between news that editors understand to be important and news that answers the wishes of their increasingly interactive and demanding readers. The challenge facing news organizations is to integrate the ecosystem of networked journalism with a diverse and engaging set of news articles. This should prove challenging in view of the significant level of fragmentation around topics, with most social networks solely focusing on two or three topics.

The debate on the effects of fragmentation and personalization due to digital media are long-running and can inform some of the issues raised by the results of our analysis. The first description of a news ecosystem dominated by customized and personalized media was presented by Negroponte (1995), who coined the term "The Daily Me" to refer to a daily newspaper tailored to individual's interests and tastes. During the 1990's and early 2000's the concept became an object of intense debate, with critics arguing that daily newspapers customized to the tastes of a person would result in greater positive feedback and reduce information choices due to the permanent reinforcement of individual's interests (Johnson 2001). The term was later associated with the customization of news feeds and links between a community of like-minded bloggers. The debate focused again on the effects of positive feedback resulting from the exposure to content that users were inclined to agree (Pariser 2011; Sunstein 2009).

Despite the contentious nature of the debate on the effects of positive feedback on digital media, the findings reported in this paper indicate that the audience of social networking sites engaging to the content published by the Times and the Guardian is slightly in favor of hard news over soft news, particularly opinion pieces, when compared to the news editors' choices. It is our understanding that this difference may be due to the particular demographics of print and social media. While both social media and tabloid readerships are on average younger, tabloid readers are on average less educated (Andersen 2003) in comparison to the audience of social networking sites, which is also more likely to be urban and politically engaged (Pew Research Center 2012b). Twitter and Facebook are especially appealing to urban adults aged 18-29, particularly Twitter, which appeals to urban dwellers more than both suburban and rural residents. Pinterest is particularly attractive to women, but also to white, young, well-educated, and wealthier people (Pew Research Center 2013).

The largely young audience of social media might also explain why news items covering economy are less suited for the user base of the social networking sites investigated in this study. Rather than picturing an utter disconnection between editors and readers, the differences between readership in social and legacy media are likely due to the younger demographics of social media users. As social media audience ages or the user base expands

and incorporates older populations, we should expect sports-loving readers and users passionate about economics to engage with social media and challenge the current distribution of shared, pinned, and tweeted news items on social networking sites. In short, we expect the demographics of social media readership to offer a possible explanation for the large volume of articles about sports and economy published by the The Guardian and The New York Times that fail to engage social media users.

On the other hand, previous research had already reported that online readership spend less time reading the sports section of a newspaper (D'Haenens et al. 2004). If this trend is not reversed, and considering social media audience also emphasizes editorials and comment pieces, we can expect a shift in the paradigm of newsworthiness towards hard news and away from the context of print media and the editorial emphasis on infotainment news. As social networks become more widely used, further research should continue to investigate whether the networked architecture of journalism is moving towards a context of editorial emphasis on comment pieces and hard news. In this hypothetical scenario, the paradigm of newsworthiness would change again and close the loop on the relationship between newsworthiness and hard news, which was originally described as having a higher level of newsworthiness (Lehman-Wilzig and Seletzky 2010).

In conclusion, we have studied in this paper the relationship between two nationwide newspapers and social media, but further research should examine if the differences are consistent across a larger set of newspapers in different linguistic and cultural settings. We have shown that social media users behave differently than the traditional readership of online newspapers, and that social networks not usually associated with news (i.e. Pinterest and StumbleUpon) play an important role in distributing information about niche topics to communities of interest. Social networks that are primarily a visual medium (i.e. Pinterest and YouTube) can specialize in providing visual and audio information in a similar fashion to broadcast television and radio. This specialization can result in greater fragmentation and boost new types of attention paid to content, further stratifying readership according to interests of like-minded groups. Whichever scenario prevails, it seems likely that social networks will play a fundamental role in shaping the networked architecture of journalism.

Acknowledgements

This research was supported by the São Paulo Research Foundation (Grants 10/06243-4 and 11/22495-6) and completed while the author was a visiting research fellow at the Department of Media and Communications of the London School of Economics and Political Science. The author is thankful to Charlie Beckett for his insightful observations and suggestions.

References

- Aitamurto, Tanja, and Seth C. Lewis. 2013. "Open Innovation in Digital Journalism: Examining the Impact of Open Apis at Four News Organizations." *New Media & Society* 15, no. 2: 314-331.
- Althaus, Scott L, and David Tewksbury. 2002. "Agenda Setting and the "New" News Patterns of Issue Importance among Readers of the Paper and Online Versions of the New York Times." *Communication Research* 29, no. 2: 180-207.
- Andersen, Robert. 2003. "Do Newspapers Enlighten Preferences? Personal Ideology, Party Choice and the Electoral Cycle: The United Kingdom, 1992-1997." *Canadian Journal of Political Science/Revue canadienne de science politique* 36, no. 03: 601-619.
- Atanassoff, Dagmar, and Silke Oehlmann. 1997. "Zur Entwicklung Der Reichweiten Der Tageszeitungen." *Bundesverband Deutscher Zeitungsverleger e. V.(Ed.), Zeitungen '97*.

- Bantimaroudis, Philemon, and Hyun Ban. 2001. "Covering the Crisis in Somalia: Framing Choices by the New York Times." edited by Stephen D. Reese, Oscar H. Gandy and August E. Grant. London: Taylor & Francis.
- Guardianr: Guardian Api Wrapper. CRAN.
- Bastos, Marco Toledo, Rodrigo Travitzki, and Cornelius Puschmann. 2012. "What Sticks with Whom? Twitter Follower-Followee Networks and News Classification." Paper presented at the The Potential of Social Media Tools and Data for Journalists in the News Media Industry - Sixth International AAAI Conference on Weblogs and Social Media, Dublin, Ireland.
- Broadgate Mainland. 2011. "Digital Trends Survey."
- Caruso, Andi. 2012. "Introducing Wikistats – What's Trending on Wikipedia Right Now." In *DataSift Blog*: DataSift.
- Christians, Clifford G., Theodore L. Glasser, Dennis McQuail, Kaarle Nordenstreng, and Robert A. White. 2009. *Normative Theories of the Media: Journalism in Democratic Societies*. University of Illinois Press.
- comScore. 2013a. "2013 Uk Digital Future in Focus." In *Future in Focus*, edited by Berit Block and Cathy McCarthy.
- . 2013b. "2013 Us Digital Future in Focus." In *Future in Focus*, edited by Andrew Lipsman, Carmela Aquino and Stephanie Flosi.
- Cowan, Matt. 2010. "Twitter Co-Founder Hopes to Create News Network." *Reuters*, 22/11/2010.
- D'Haenens, Leen, Nicholas Jankowski, and Ard Heuvelman. 2004. "News in Online and Print Newspapers: Differences in Reader Consumption and Recall." *New Media & Society* 6, no. 3: 363-382.
- Dans, Enrique. 2000. "Internet Newspapers: Are Some More Equal Than Others?" *International Journal on Media Management* 2, no. 1: 4-13.
- Dow, P Alex, Lada A Adamic, and Adrien Friggeri. 2013. "The Anatomy of Large Facebook Cascades." Paper presented at the 7th International AAAI Conference on Weblogs and Social Media (ICWSM13).
- Dutton, William, and Grant Blank. 2011. "Next Generation Users: The Internet in Britain."
- Franklin, Bob. 2008. "The Future of Newspapers." *Journalism Practice* 2, no. 3: 306-317.
- Galtung, Johan, and Mari Holmboe Ruge. 1965. "The Structure of Foreign News: The Presentation of the Congo, Cuba and Cyprus Crises in Four Norwegian Newspapers." *Journal of Peace Research* 2, no. 1: 64-90.
- Gasher, Mike, and Sandra Gabriele. 2004. "Increasing Circulation? A Comparative News-Flow Study of the Montreal Gazette's Hard-Copy and on-Line Editions." *Journalism Studies* 5, no. 3: 311-323.
- George, Lisa M. 2008. "The Internet and the Market for Daily Newspapers." *Journal of Economic Analysis & Policy* 8, no. 1.
- Greer, Jennifer, and Donica Mensing. 2006. "The Evolution of Online Newspapers: A Longitudinal Content Analysis, 1997-2003." *Internet newspapers: The making of a mainstream medium*: 13-32.
- Groth, Otto. 1928. *Die Zeitung; Ein System Der Zeitungskunde (Journalistik)*. Mannheim: J. Bensheimer.
- Harcup, Tony, and Deirdre O'Neill. 2001. "What Is News? Galtung and Ruge Revisited." *Journalism Studies* 2, no. 2: 276-280.
- Hart, Jim A. 1966. "Foreign News in U. S. And English Daily Newspapers: A Comparison." *Journalism & Mass Communication Quarterly* 43, no. 3: 443-448.
- Hopple, Gerald W. 1982. "International News Coverage in Two Elite Newspapers." *Journal of Communication* 32, no. 1: 61-74.
- Hume, Andy. 2013. "Breaking News and Breaking Software." In *SyncConf*.

- Johnson, Steven. 2001. *Emergence: The Connected Lives of Ants, Brains, Cities, and Software*. Nova York: Scribner.
- Johnson, Thomas J., and Barbara K. Kayer. 2004. "Wag the Blog: How Reliance on Traditional Media and the Internet Influence Credibility Perceptions of Weblogs among Blog Users." *Journalism and Mass Communication Quarterly* 81, no. 3: 622-642.
- Kamerer, David, and Bonnie Bressers. 1998. "Online Newspapers: A Trend Study of News Content and Technical Features." Paper presented at the 81th annual meeting of the Association for Education in Journalism and Mass Communication, Baltimore, MD.
- Klein, Malcolm W, and Nathan Maccoby. 1954. "Newspaper Objectivity in the 1952 Campaign." *Journalism Quarterly* 31, no. 3: 285-296.
- Koszyk, Kurt, and Jürgen Prause. 1987. "Regionalzeitungen Aus Dem Ruhrgebiet Im Bundestagswahlkampf 1986/87. Eine Themen-Und Tendenzanalyse." *Wahlen und Wähler. Analysen aus Anlaß der Bundestagswahl*: 619-646.
- Lasorsa, Dominic L., Seth C. Lewis, and Avery E. Holton. 2011. "Normalizing Twitter: Journalism Practice in an Emerging Communication Space." *Journalism Studies* 13, no. 1: 19-36.
- Lehman-Wilzig, Sam N., and Michal Seletzky. 2010. "Hard News, Soft News, 'General' News: The Necessity and Utility of an Intermediate Classification." *Journalism* 11, no. 1: 37-56.
- McCombs, Maxwell E., and Donald L. Shaw. 1972. "The Agenda-Setting Function of Mass Media." *Public Opinion Quarterly* 36, no. 2: 176-187.
- Negroponte, Nicholas. 1995. *Being Digital*. New York: Vintage Books.
- Neuberger, Christoph, Jan Tonnemacher, Matthias Biebl, and André Duck. 1998. "Online - the Future of Newspapers? Germany's Dailies on the World Wide Web." *Journal of Computer-Mediated Communication* 4, no. 1: 0-0.
- Newman, Nic. 2009. "The Rise of Social Media and Its Impact on Mainstream Journalism: A Study of How Newspapers and Broadcasters in the Uk and Us Are Responding to a Wave of Participatory Social Media, and a Historic Shift in Control Towards Individual Consumers." Oxford: Reuters Institute for the Study of Journalism.
- Pariser, Eli. 2011. *The Filter Bubble: What the Internet Is Hiding from You*. Penguin Books Limited.
- Pew Research Center. 2012a. "Changing News Landscape, Even Television Is Vulnerable. Trends in News Consumption: 1991-2012." Pew Internet & American Life Project.
- . 2013. "The Demographics of Social Media Users, 2012." edited by Maeve Duggan and Joanna Brenner: Pew Research Center's Internet & American Life Project.
- . 2012b. "Social Media and Political Engagement." edited by Lee Rainie, Aaron Smith, Kay Lehman Schlozman, Henry Brady and Sidney Verba: Pew Internet & American Life Project.
- Quandt, Thorsten. 2008. "(No) News on the World Wide Web?". *Journalism Studies* 9, no. 5: 717-738.
- R: A Language and Environment for Statistical Computing. Version 2.15.1, Vienna.
- Schudson, Michael. 2011. *The Sociology of News*. Contemporary Societies. 2nd ed. New York: W.W. Norton & Company.
- Scott, David K., and Robert H. Gobetz. 1992. "Hard News/Soft News Content of the National Broadcast Networks, 1972–1987." *Journalism & Mass Communication Quarterly* 69, no. 2: 406-412.
- Shiels, Maggie. 2011. "Twitter Co-Founder Jack Dorsey Rejoins Company." *BBC News*, 28/03/2011.
- Singer, Jane B. 2007. "Contested Autonomy." *Journalism Studies* 8, no. 1: 79-95.
- Sunstein, Cass. 2009. *Republic.Com 2.0*. Princeton: Princeton University Press.

- Tunstall, Jeremy. 1971. *Journalists at Work: Specialist Correspondents: Their News Organizations, News Sources, and Competitor-Colleagues*. Communication and Society. London,: Constable.
- Twitter API Documentation. 2012. "Issue #48: Issues with Tweet Button Counting."
- Whittaker, Zack. 2012. "Facebook Hits 1 Billion Active User Milestone." *CNET*, 04/10/2012.
- Wickham, Hadley. 2009. *Ggplot2: Elegant Graphics for Data Analysis*. Springer Publishing Company, Incorporated.
- Wolff, Michael. 2013. "Twitter Is About to Change the News Media – Again." *The Guardian*.

Contact details

Marco Toledo Bastos
HASTAC, Duke University
114 S. Buchanan Blvd, Bay 5
Box 90403 / Durham, NC 27708
<http://marco.toledobastos.com/>

TABLE 1

	Twitter		Facebook			
	Links	Retweets	Links	Likes	Shares	Comments
NY Times	4,264	167,189	4750	634,633	2,115,513	418,532
Guardian	5,050	254,506	4501	245,890	856,149	186,452
	Newspaper		StumbleUpon	Pinterest	Delicious	Google+
	Items	Word count	Hits	Hits	Hits	Hits
NY Times	11,607	5,641,129	150,466	3,561	1,216	5,626
Guardian	5,222	3,567,860	398,401	4,390	1,738	11,532

Table 1 Number of news articles per newspaper and absolute number of hits on social media platforms

FIGURE 1

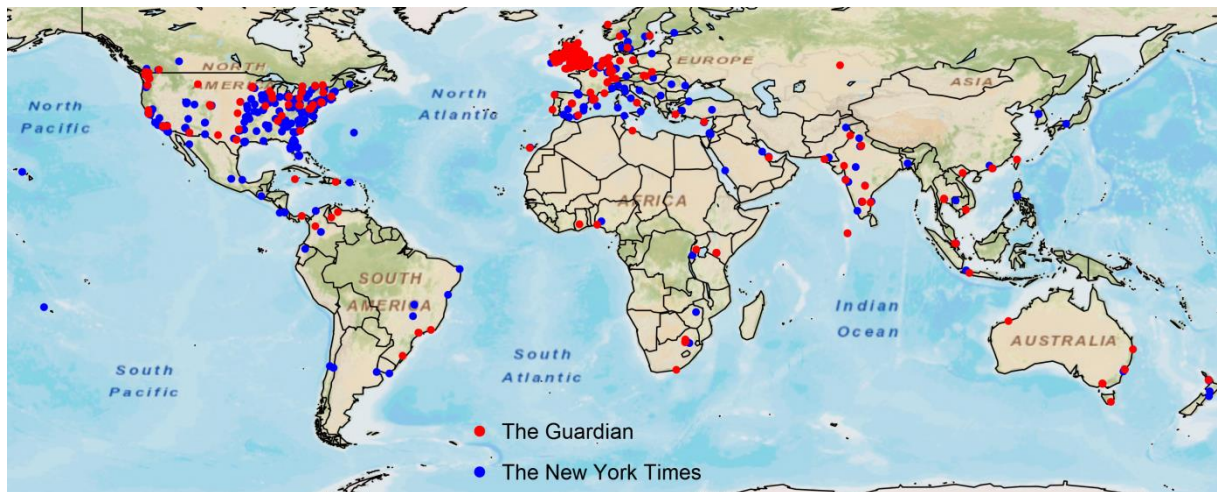


Figure 1 Geolocation of tweets with links to The Guardian and The New York Times (sample with 3% of total tweets)

FIGURE 2

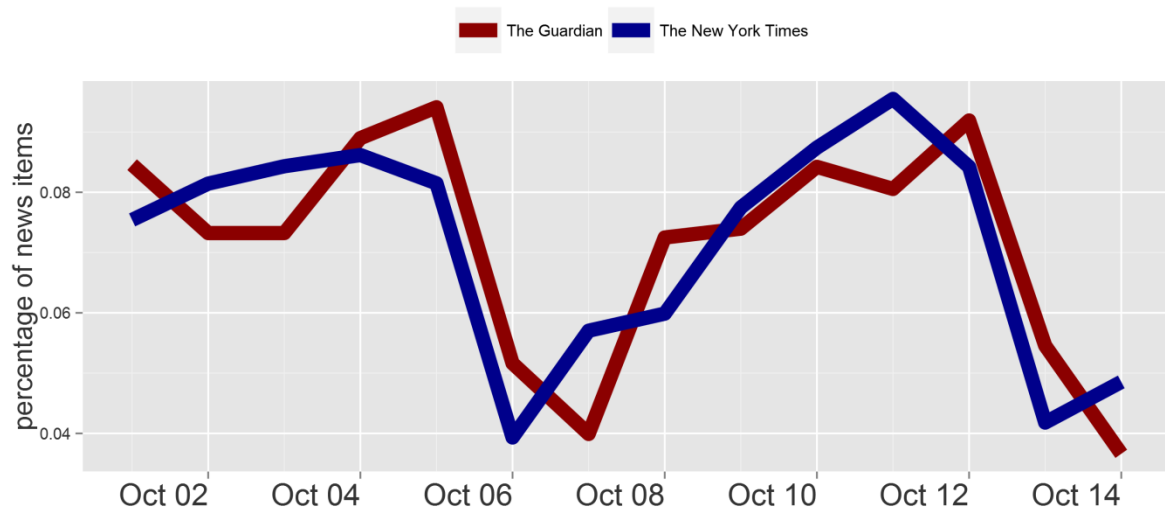


Figure 2 Time series of articles published by The New York Times and The Guardian during 1-14 October, 2012

TABLE 2

	Word count	Tweets	Retweets	Comments	Shares	Likes	Delicious
Guardian	0,96	0,89	0,86	0,79	0,78	0,73	0,63
NY Times	0,97	0,68	0,65	NA	NA	NA	NA

Table 2 Correlation between the volume of news items per section in newspapers and social networks (p<0.001)

FIGURE 3

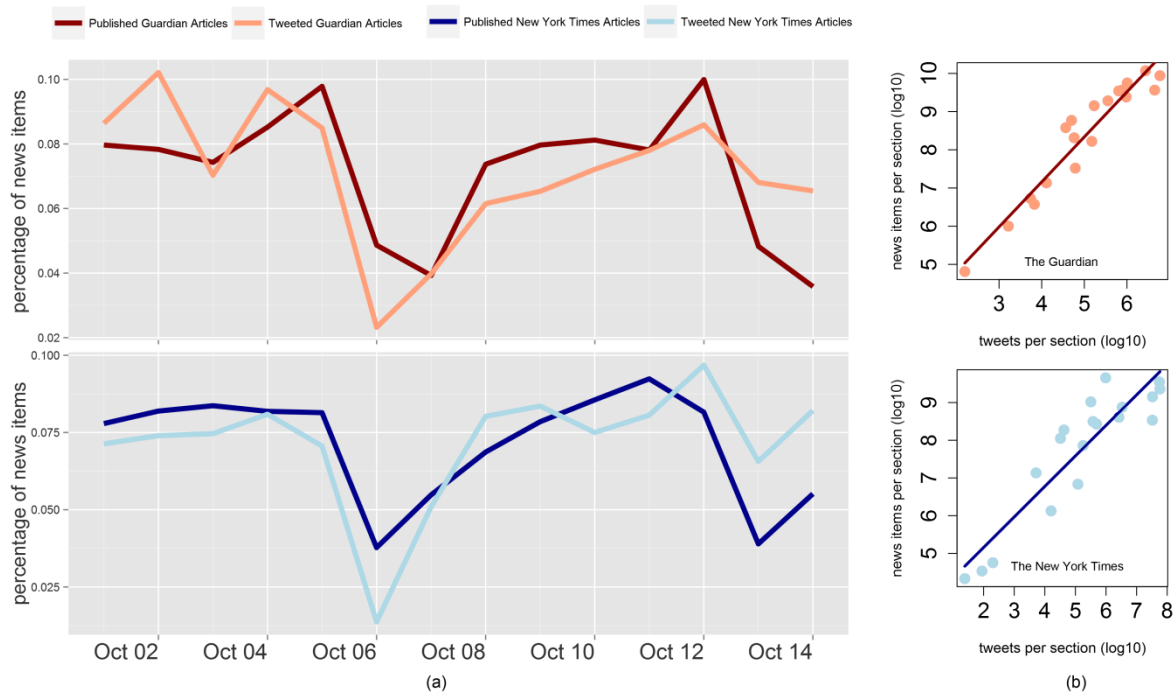


Figure 3 Time-series of published and tweets articles of The Guardian and The New York Times

FIGURE 4

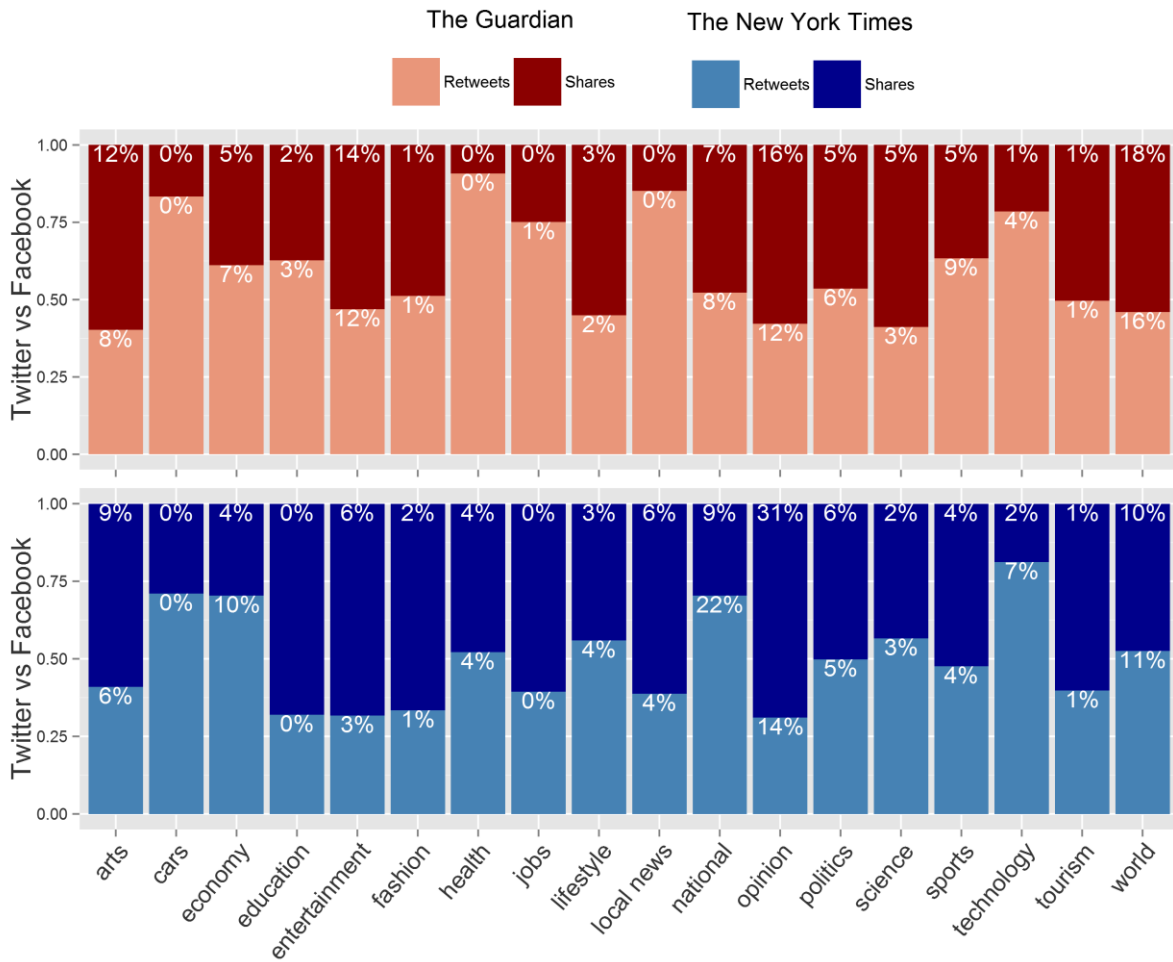


Figure 4 Comparison between Twitter and Facebook in terms of retweeted and shared items per news sections

FIGURE 5

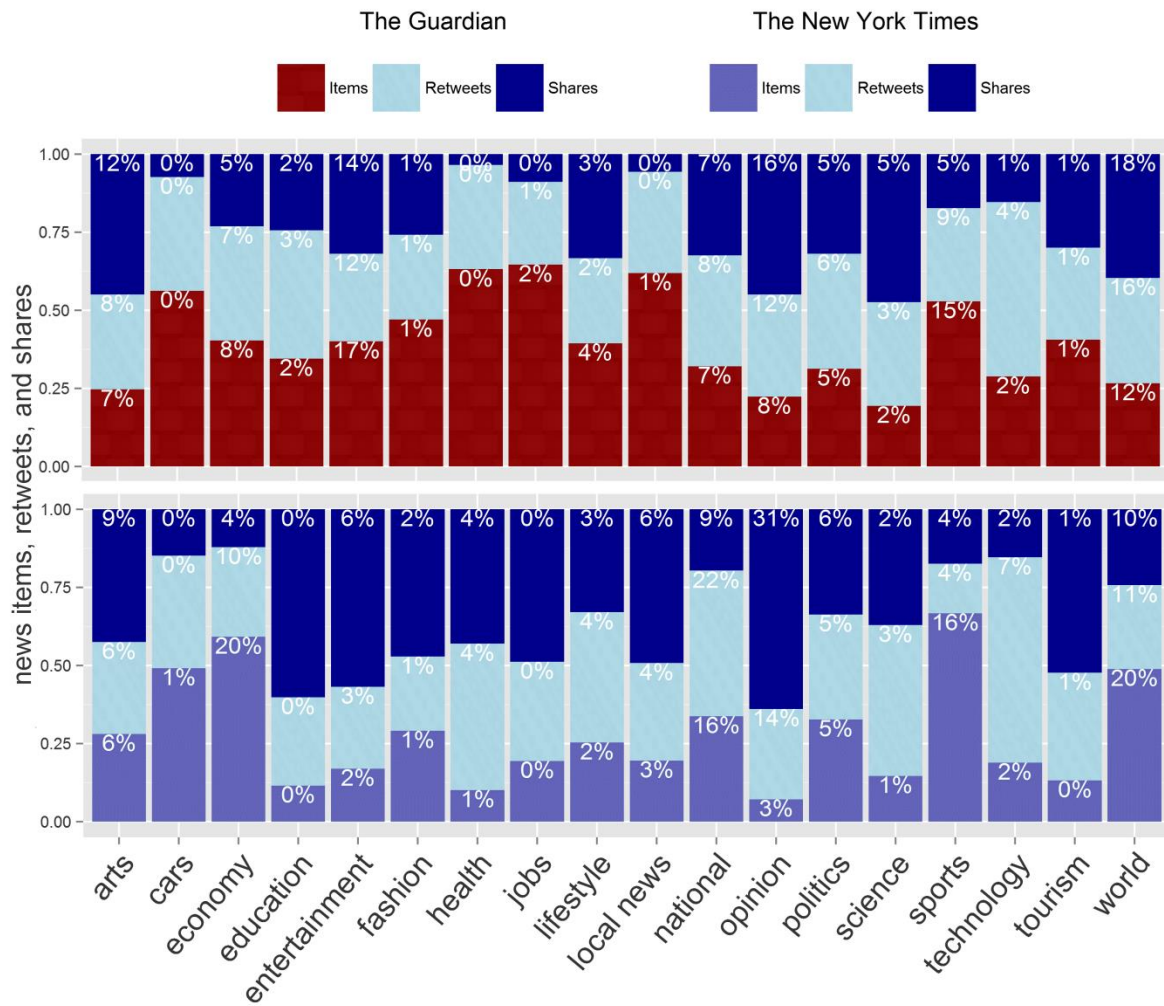


Figure 5 Percentage of shares, retweets, and news articles per section (bars of the same color sum up to 100%)

FIGURE 6

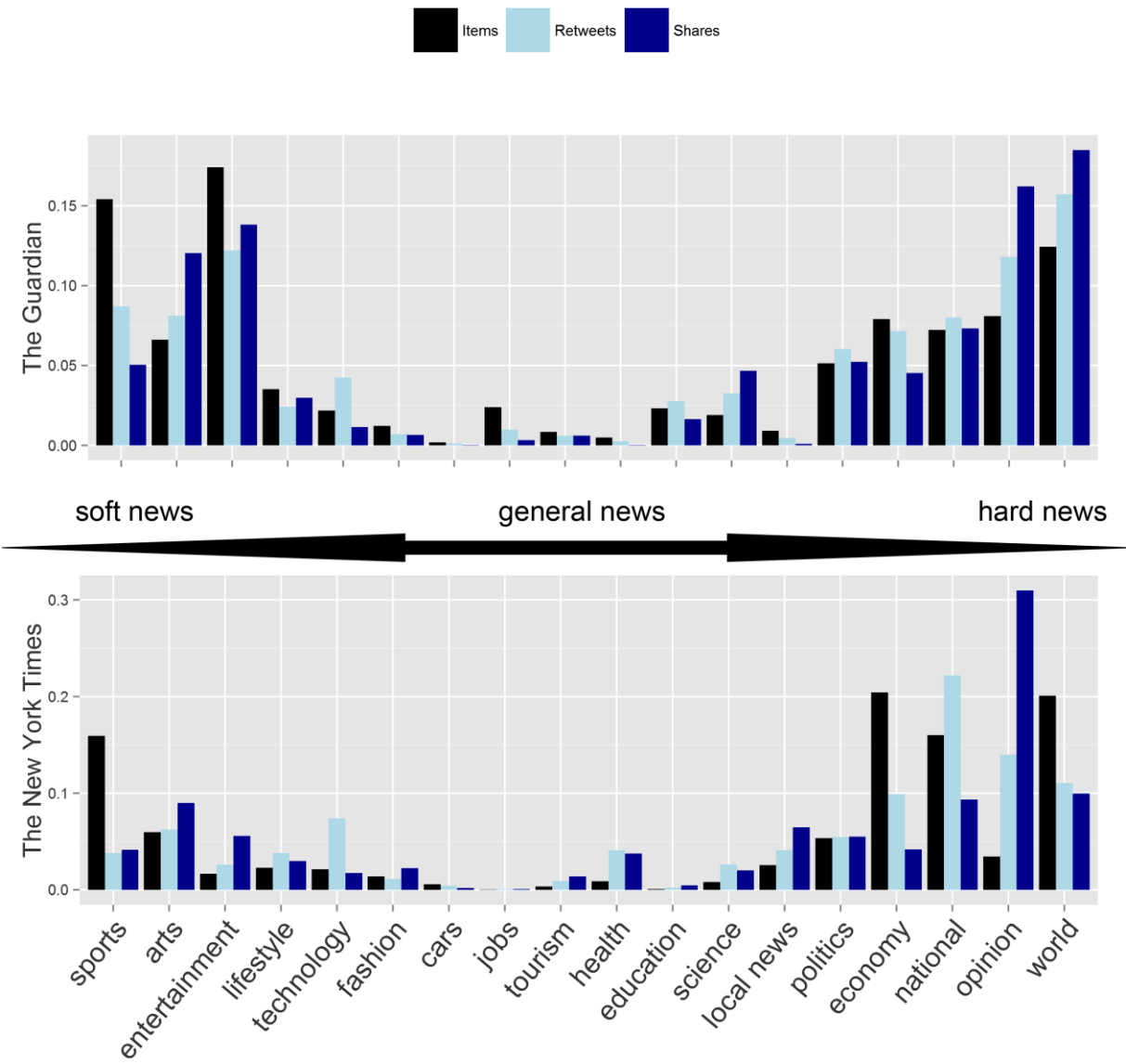


Figure 6 Distribution of news articles by soft, general, and hard news on news outlets, Twitter and Facebook

FIGURE 7

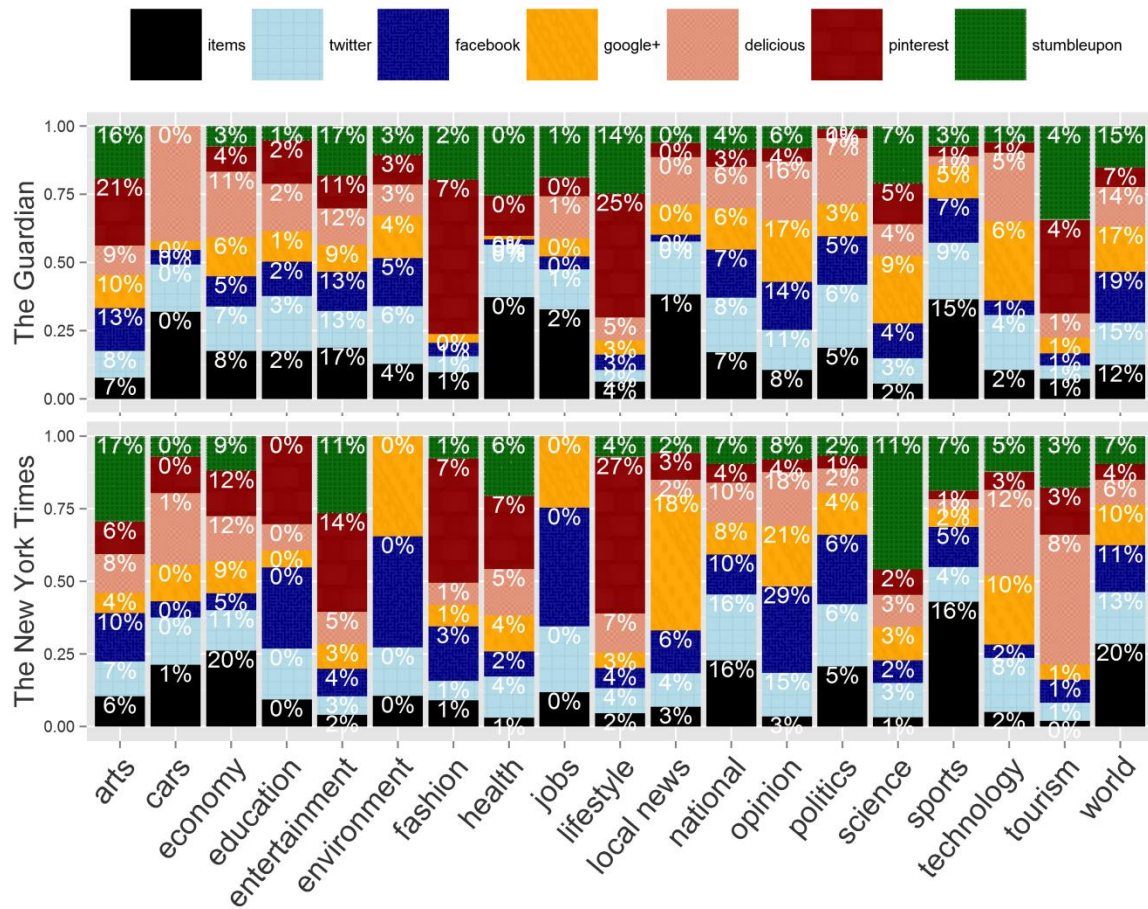


Figure 7 Ratio of articles per news section on the Times and the Guardian compared to the ration of hits per section on Twitter, Facebook, Google+, Delicious, Pinterest, and StumbleUpon (bars of the same color sum up to 100%)